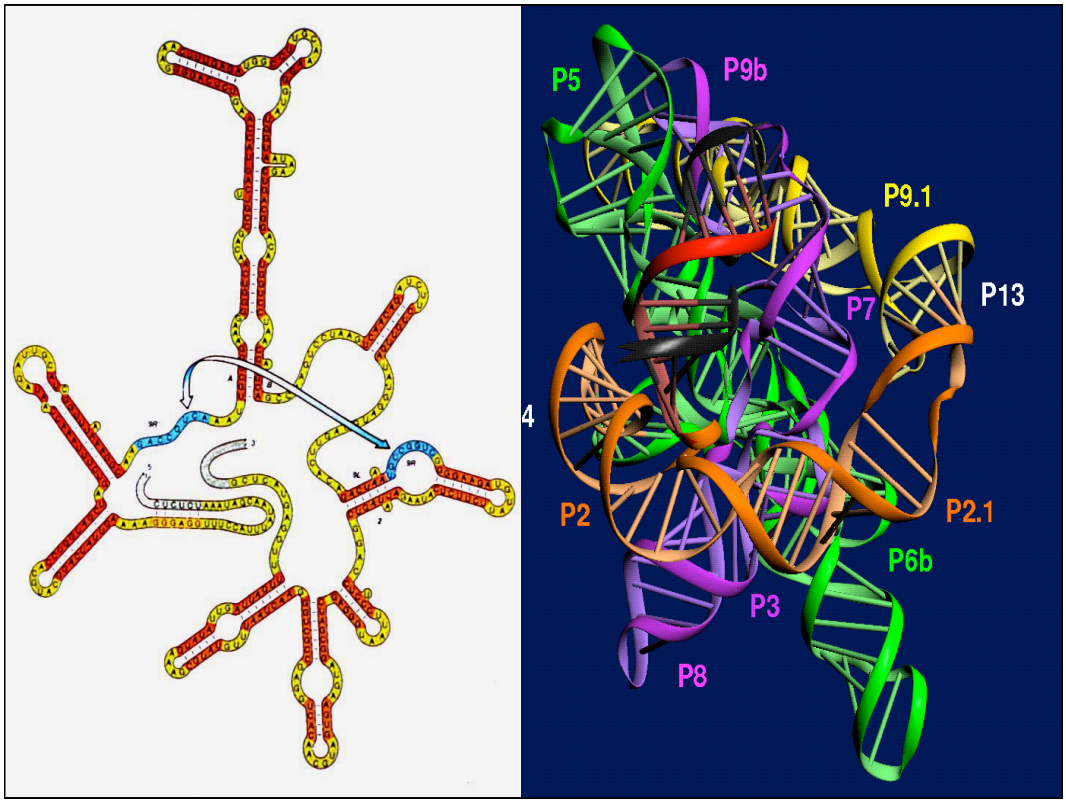
Les modules récurrents de l'ARN
et leur diversité

E. Westhof

2010

Institut de Biologie Moléculaire et Cellulaire

Some insights into RNA Architecture



RNA Folding

is modular & hierarchical

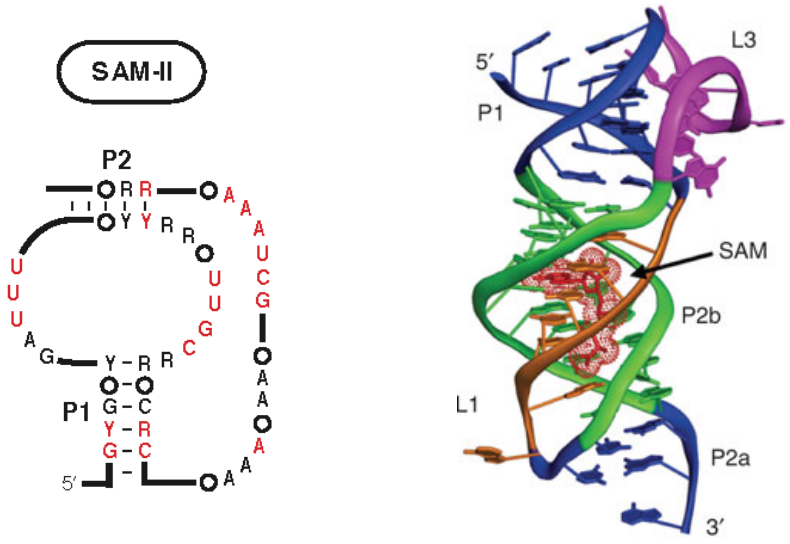
1/ Formation of helices, hairpins

2/ End-to-end stacking, parallel packing of helices

3/ Formation of tertiary contacts (helix/loop, loop/loop,...)

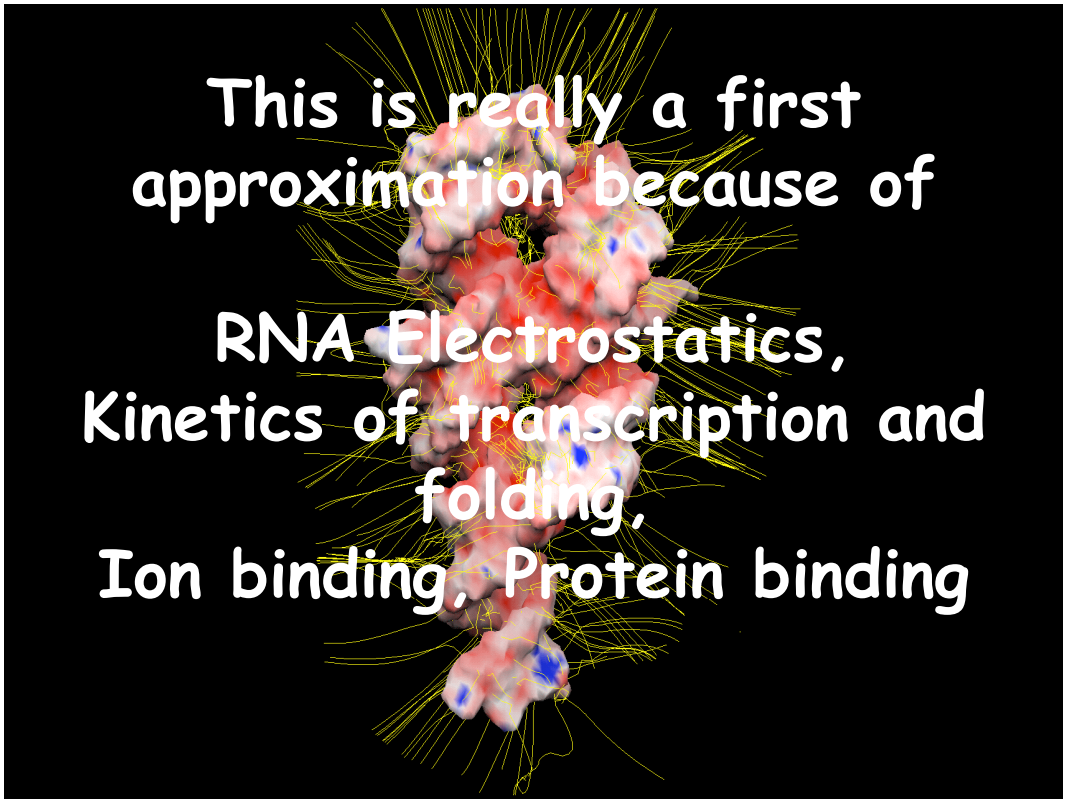
RNA Folding

Stacking of bases
and
base-base H-bonding
are both key for RNA architecture



Structure of the SAM-II riboswitch bound to S-adenosylmethionine

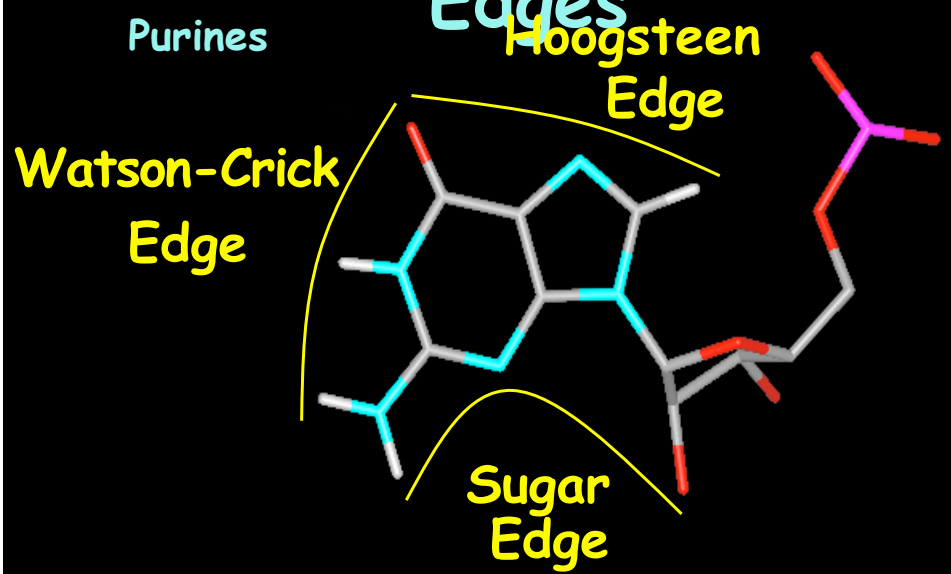
Gilbert, Rambo, Van Tyne, & Batey, Nature Struct & Mol Biol 15, 177 (2008)



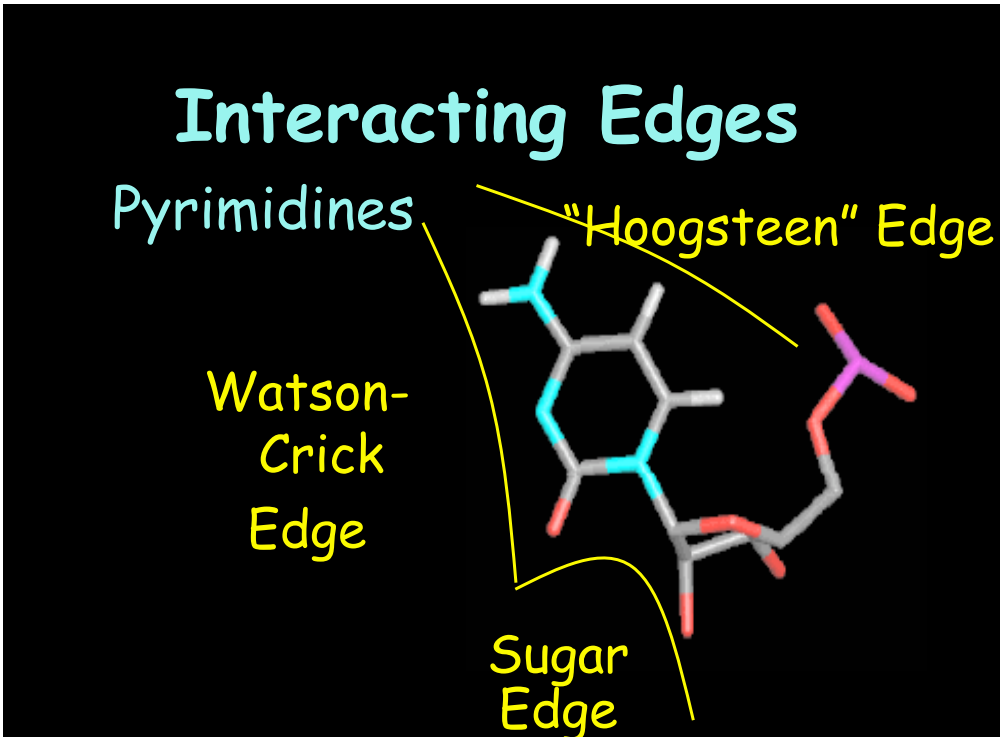
An ontology for base pairing

- Watson-Crick pairs (digital mode)
  - Form standard RNA helices
  - Define the 2D structure
- Non-Watson-Crick pairs (analog mode)
  - Form 3D RNA modules
  - Responsible for RNA-RNA recognition & 3D fold

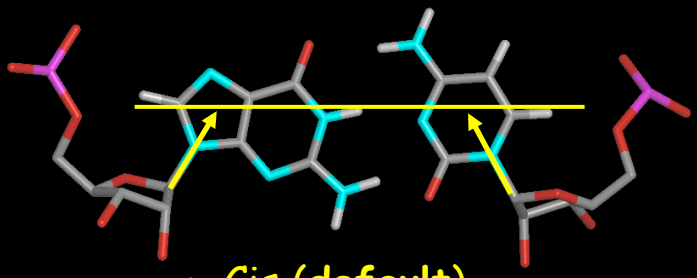
Three Interacting Edges



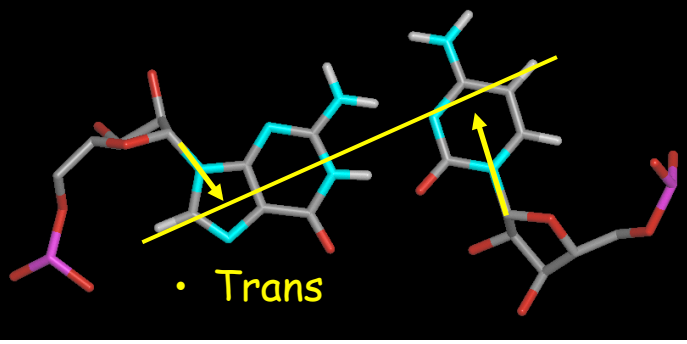
Interacting Edges



Glycosidic Bond Orientation



- Cis (**default**)

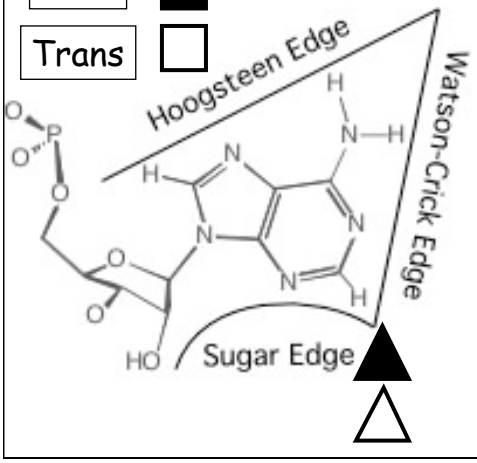


- Trans

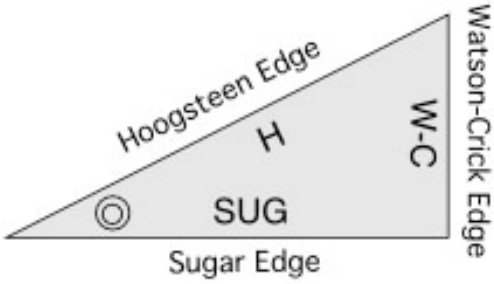
Each base has three edges and each edge carries various H-bonding sites

Cis ■

Trans □



● ○



Edge-to-Edge Pairing Families

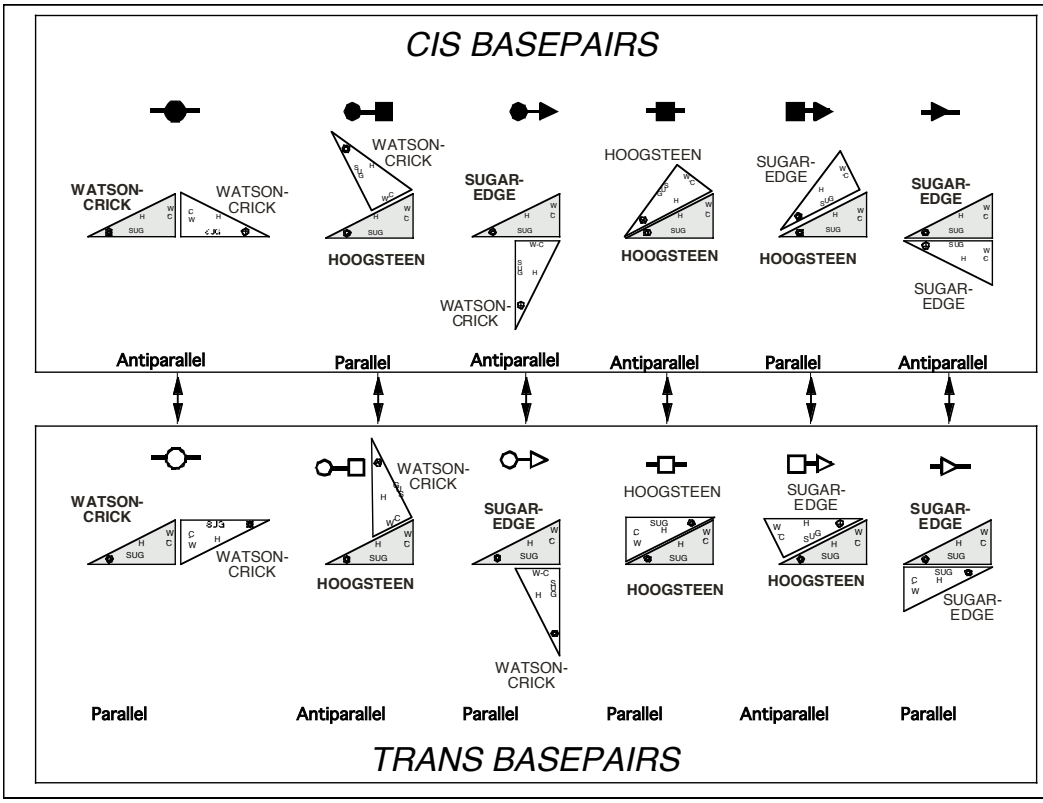
Watson-Crick
Hoogsteen
Sugar-edge

} { Watson-Crick
Hoogsteen
Sugar-edge

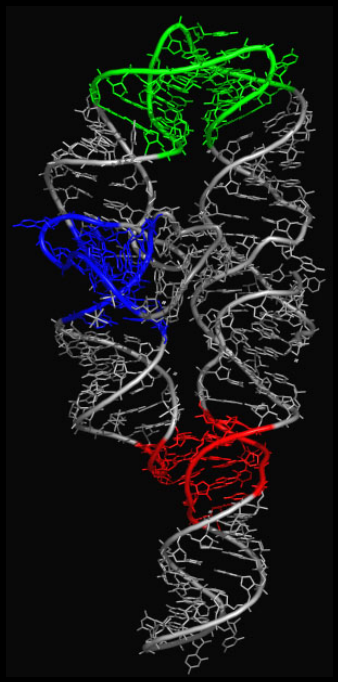
} { Cis
Trans

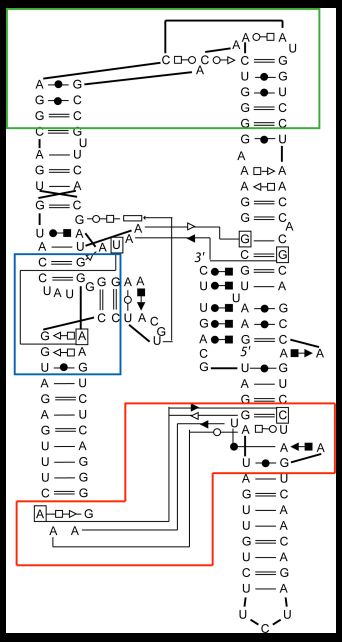
= 12 Basic Families

RNA **7**, 499 (2001)



Each structure can be parsed in modules that are recurrent in various RNAs





Bases forming base pairs
cWW base pairs and no non-cWW base pair 52%
cWW base pairs and at least one non-cWW base pairs 7%
At least one non-cWW base pairs and no cWW base pairs 20%

Bases forming other interactions (no base pairing)
Base-stacking and base-phosphate interaction 13%
Base-stacking only 3%
Base-phosphate only 1%

Bases forming no RNA-RNA interactions 4%

Total 100%

Those recurrent non-Watson-Crick base paired modules

-produce helical variations and bending within helical domains

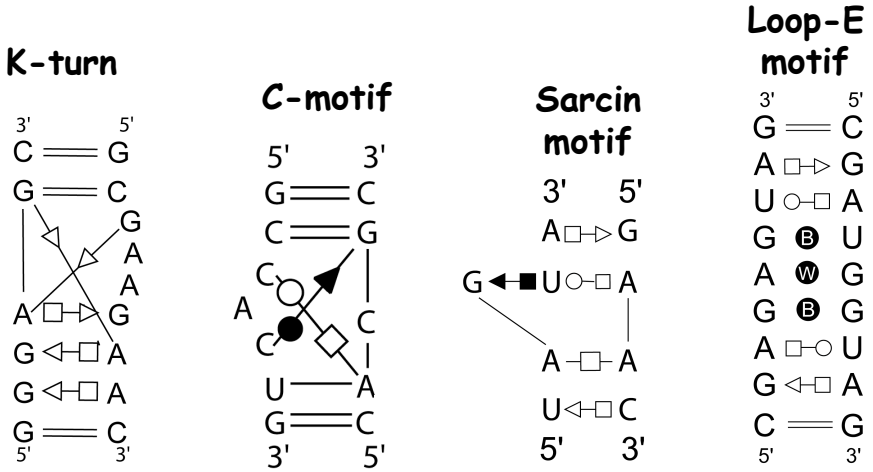
-organize locally junctions and helical domains

-participate in long-range RNA-RNA interactions between domains

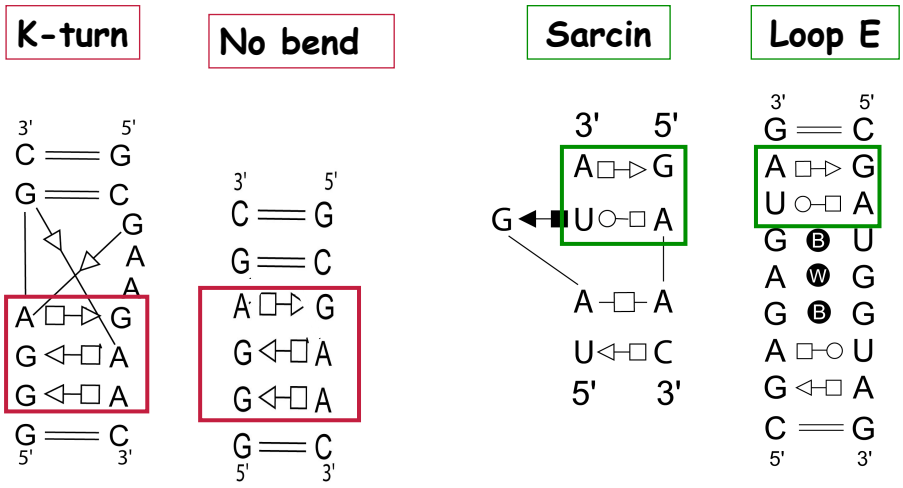
-bind small ligands

- participate in RNA-protein interactions

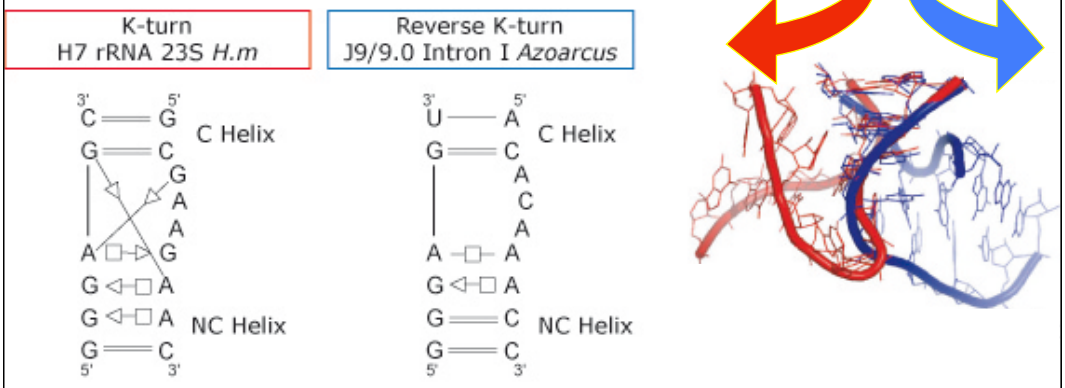
A MODULE is an ensemble of ordered non-Watson-Crick base pairs.



Modules within modules.

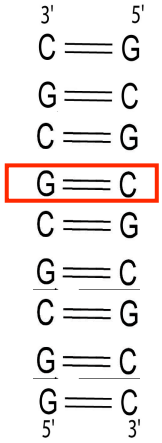


Depending on the non-WC pairs the turn is either to the left or to the right

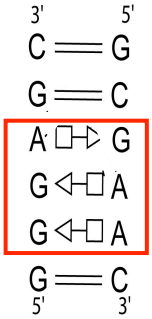


Basic units are different

2D structure



3D structure

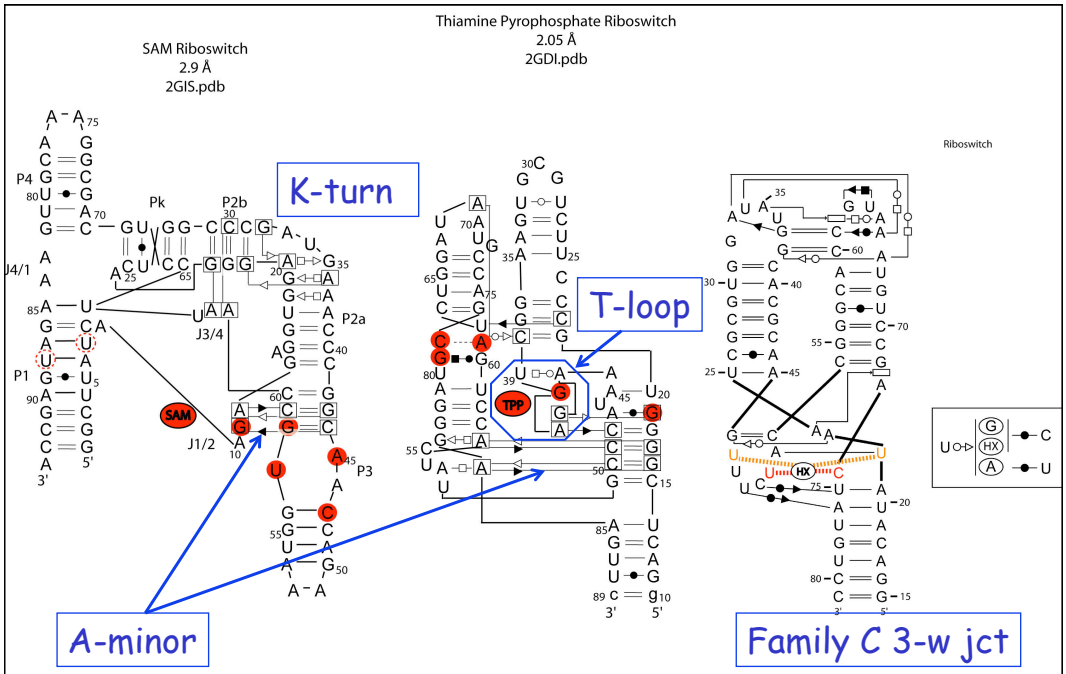


Nearest neighbour approx

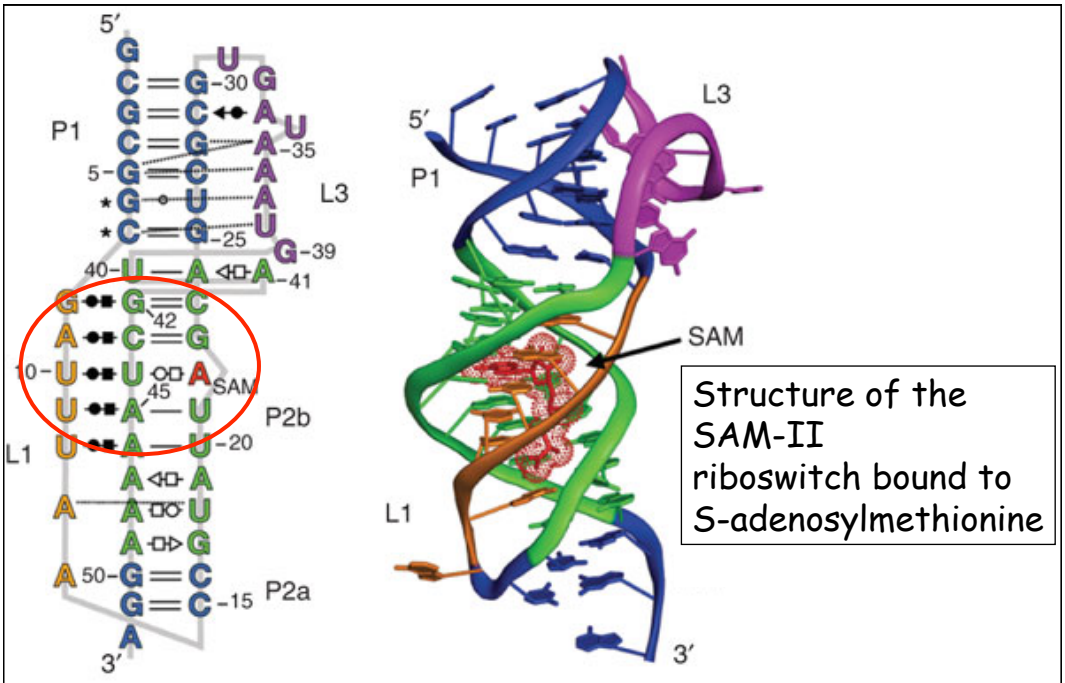
Monovalents

Set of non-WC

Divalents



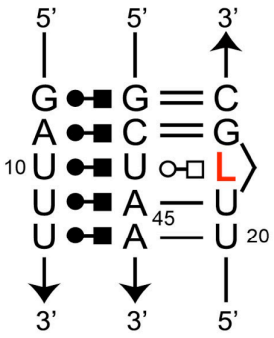
The modules are recurrent



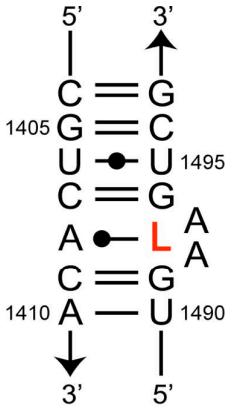
Structure of the SAM-II riboswitch bound to S-adenosylmethionine

Gilbert, Rambo, Van Tyne, & Batey, Nature Struct & Mol Biol 15, 177 (2008)

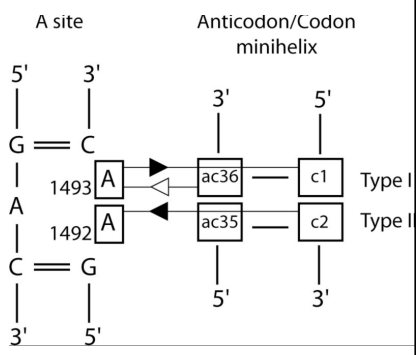
Annotations



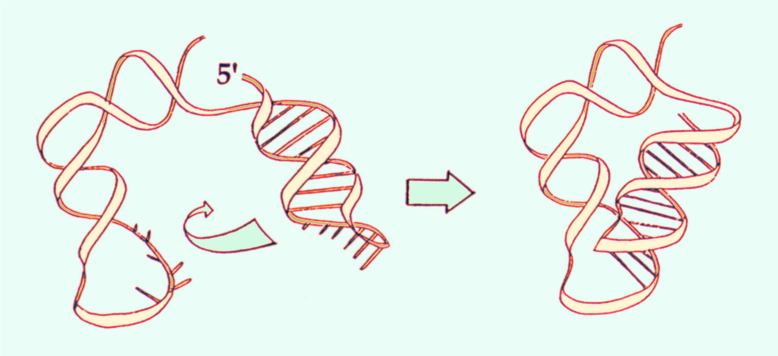
SAM-II riboswitch
(Base pair)



Bacterial A site
(Pseudo pair)



The assembly Modules



Those recurrent non-Watson-Crick base paired modules

-produce helical variations and bending within helical domains

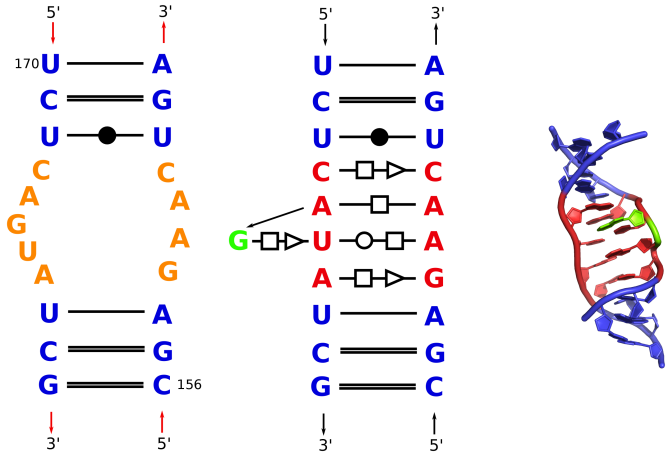
-organize locally junctions and helical domains

-participate in long-range RNA-RNA interactions between domains

-bind small ligands

- participate in RNA-protein interactions

Organization of an internal loop



CA
G • A
C - G
C - G
G • U
G • U
Y • R
A • A
U • A
G
A • A
G
G - C
G • U
G - C
5' 3'

890 -

885 - 912

Predicted

The switch helix
H27

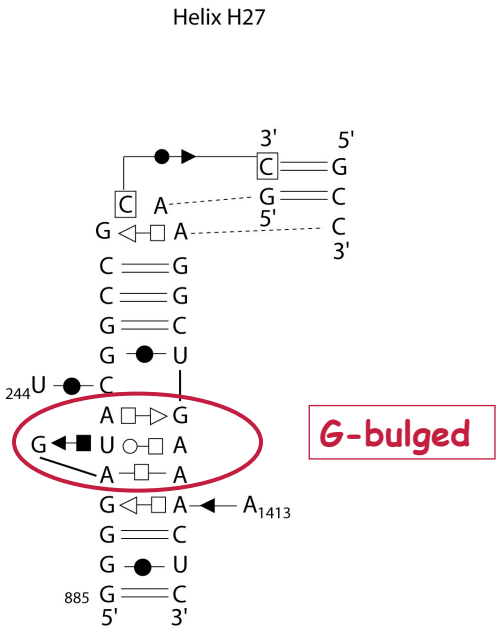
CA
G • A
C - G
C - G
G • U
G • U
Y
A • R
U • A
G
A • A
G • A
G - C
G • U
G - C
5' 3'

885 - 912

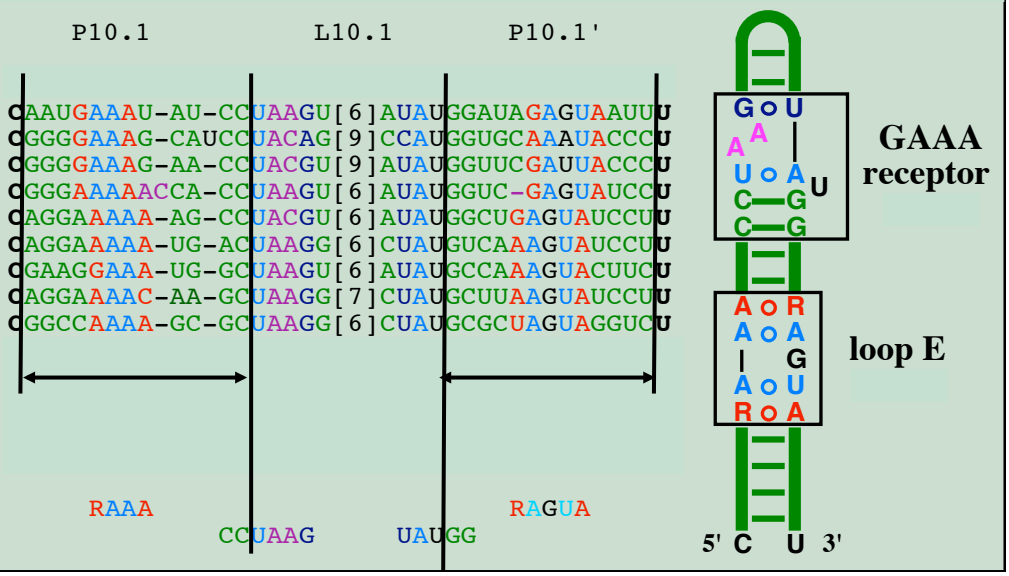
Observed

Participation in long-range RNA-RNA interactions between domains

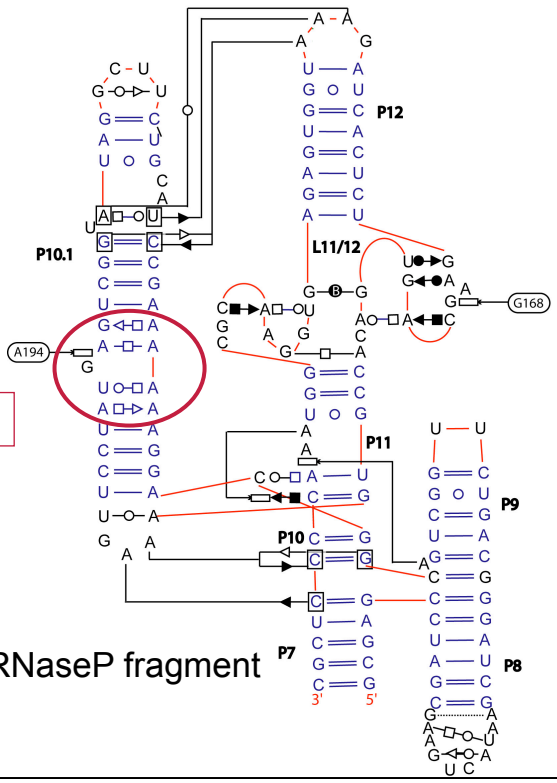
16S rRNA



Relative distances between modules



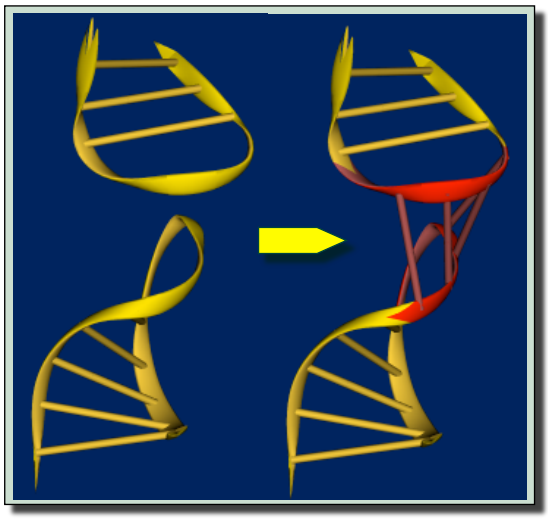
Participation in long-range RNA-RNA interactions between domains



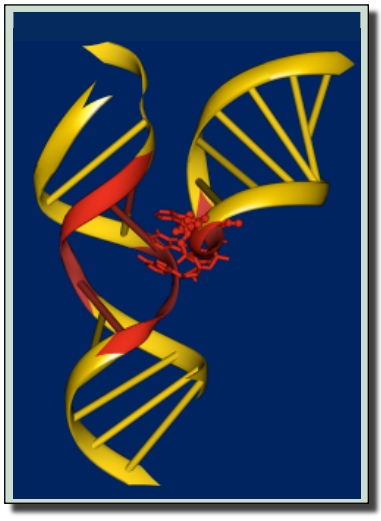
G-bulged

The key RNA self-assembly modules

-loop / loop -



-A-minor modules -



The self-assembly modules

Sequence-based : Watson-Crick base pairs

- guide sequence,
- loop-loop,
- pseudoknots

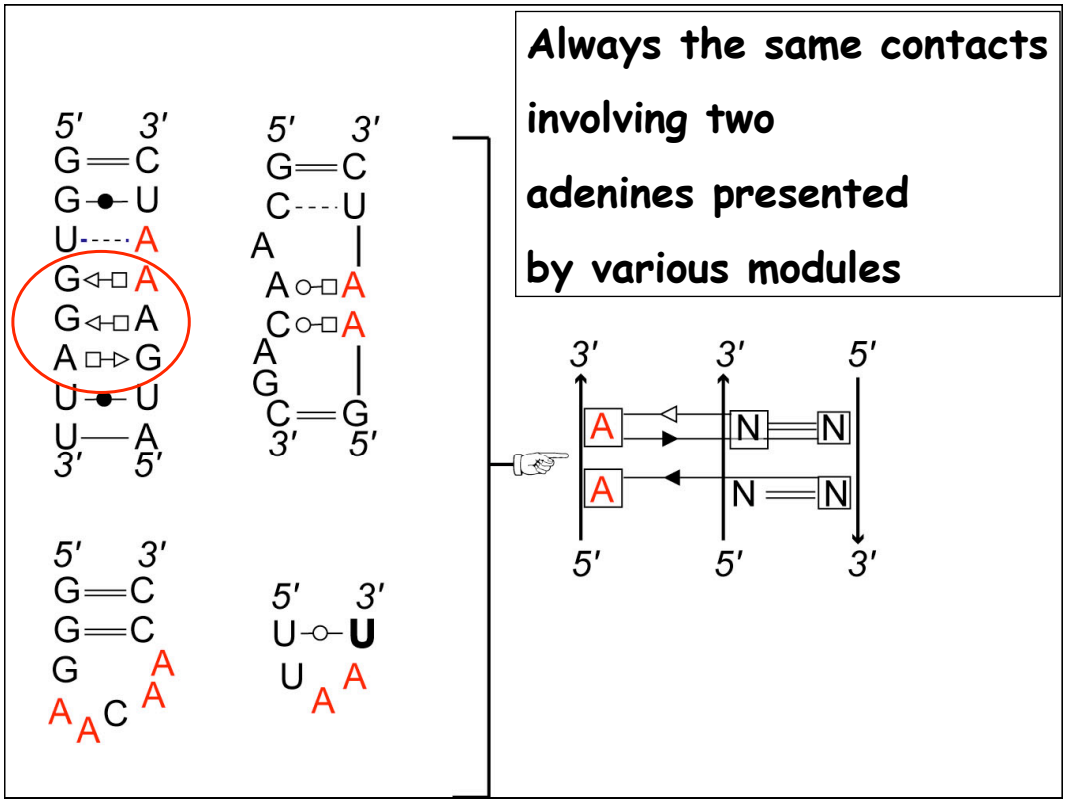
Backbone-based : A-minor interactions

Non-specific : between sugar-phosphate

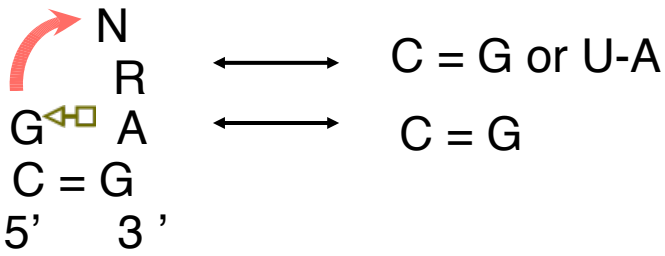
Ribose zippers (O2'...O2'; O2'...N3(R)/O2(Y)

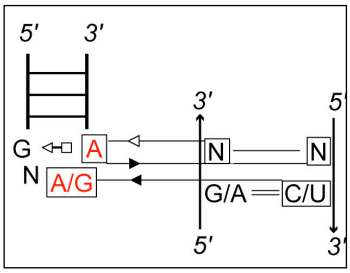
Specific on one side : 5'AA3' ... WC pairs

Highly specific : GNRA...WC pairs/11nt-modif



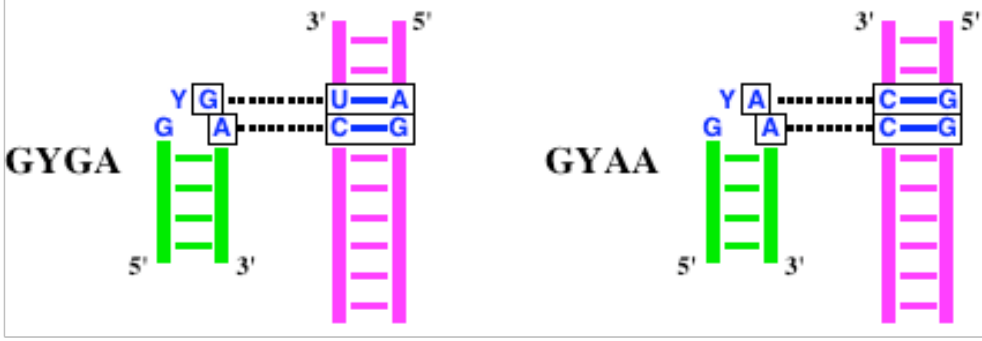
GNRA - helix/receptors





GYGA CCAGUUGGCC[ ]GGGGGCGACCCC[ ]GGCCAACUGG
AGAGUCGAUU[ ]GGGCGUGAGCCU[ ]GAUCGACUCU

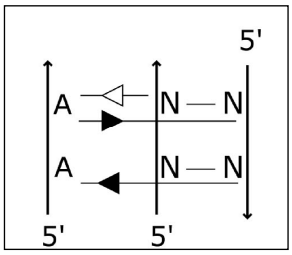
GYAA CAGGAUAGGG[ ]CGGGGCAACCCG[ ]CCCUAUCCUG
CGGGACGAGG[ ]CGGGGUAACCCG[ ]CCUCGUCCCG



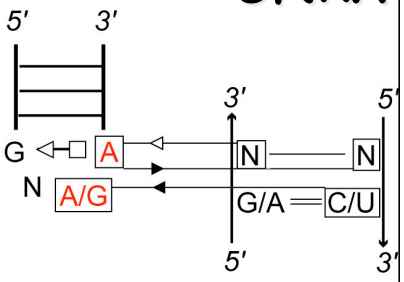
Variants of GNRA type of contacts

Most common

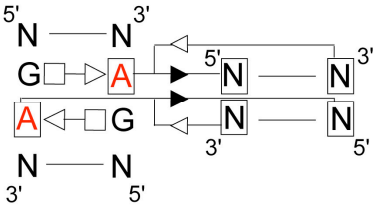
TypeI/TypeI



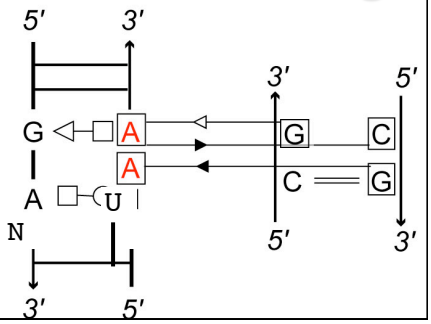
GNRA



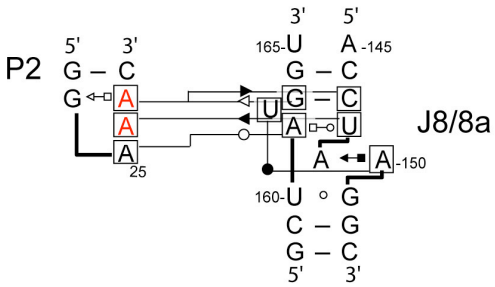
GA tandem



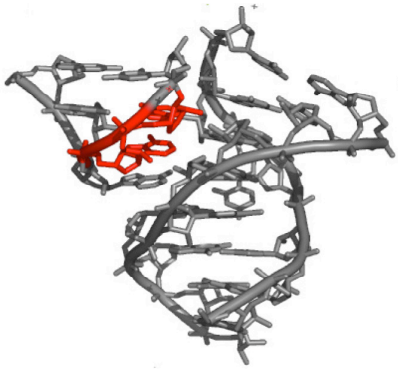
A-rich bulge

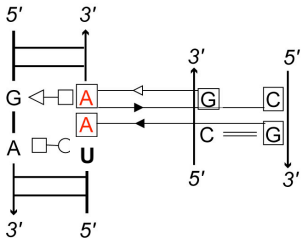


A more complex variant: The 11nt-motif interacts with *GAAA* tetraloop

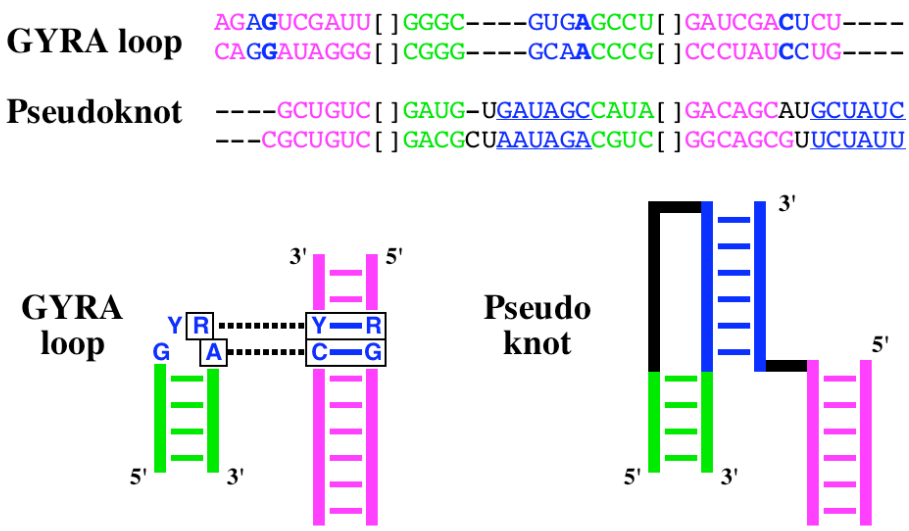


J8/8a

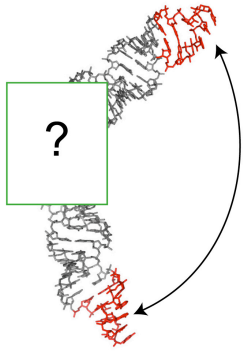




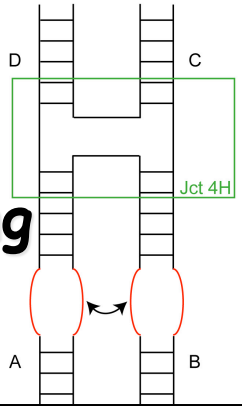
Motif Swap between GNRA/helix and loop-loop motifs

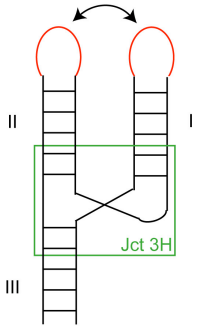


Topology is key to guarantee specificity despite neutrality

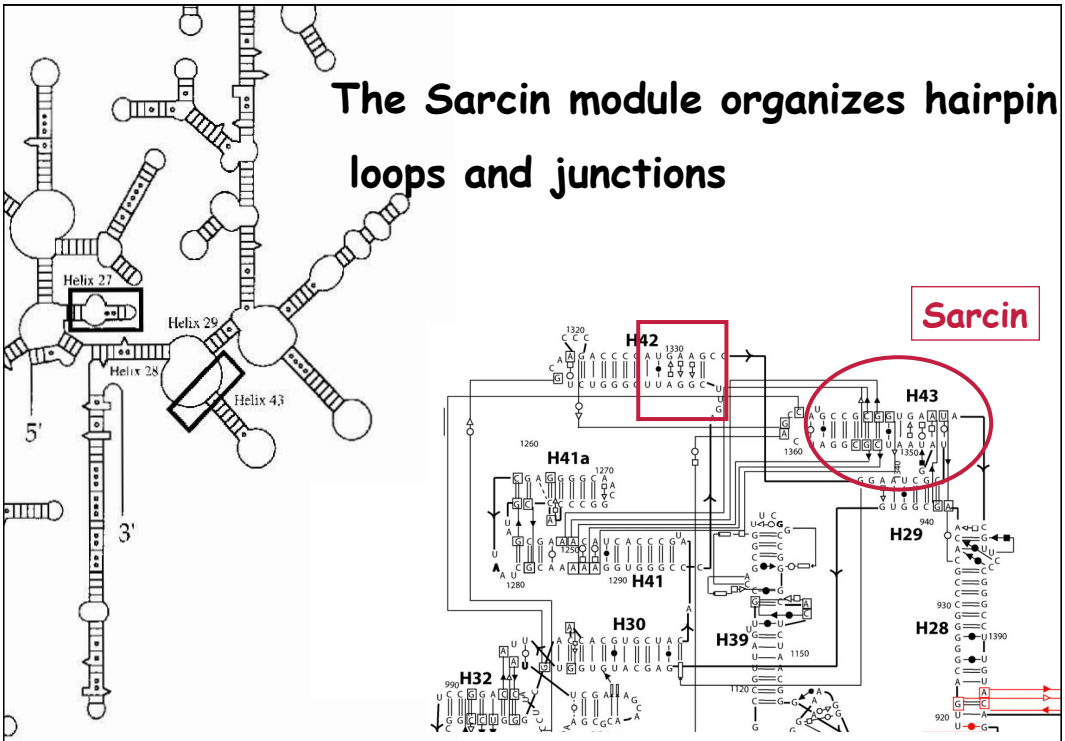


Co-axial stacking & junctions between helices

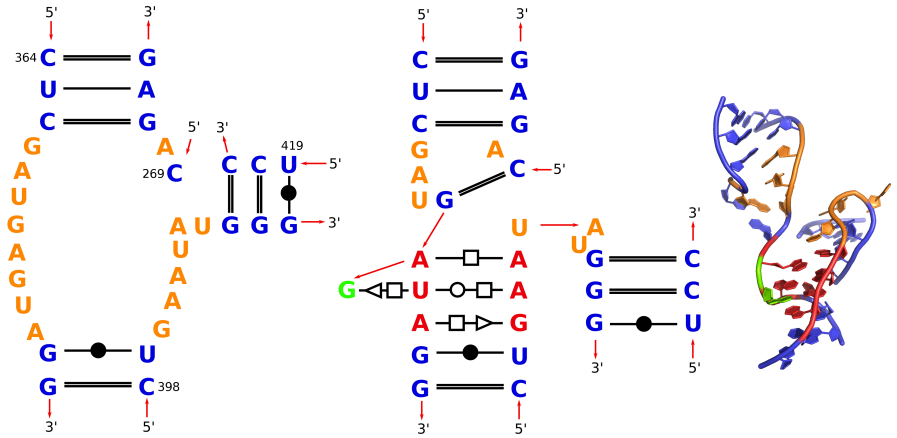


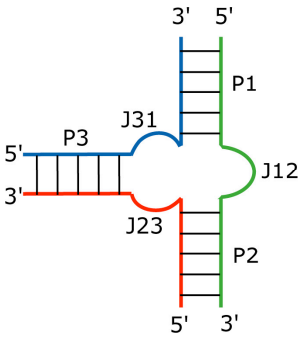


The Sarcin module organizes hairpin loops and junctions

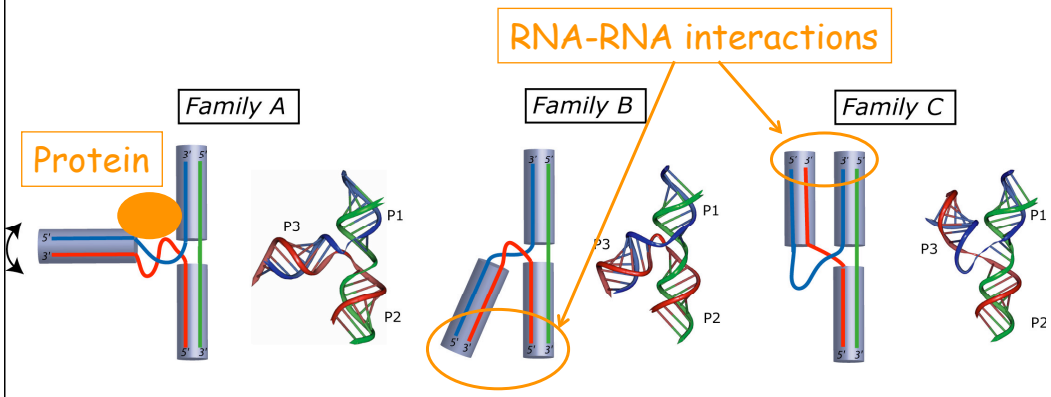


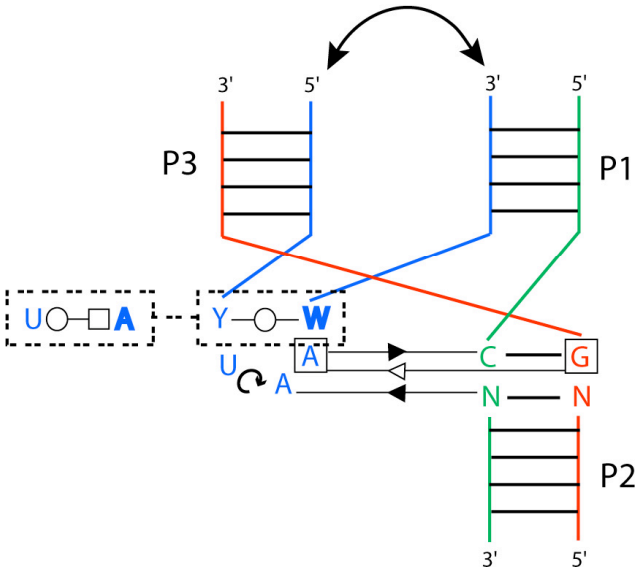
Organization of a three-way junction



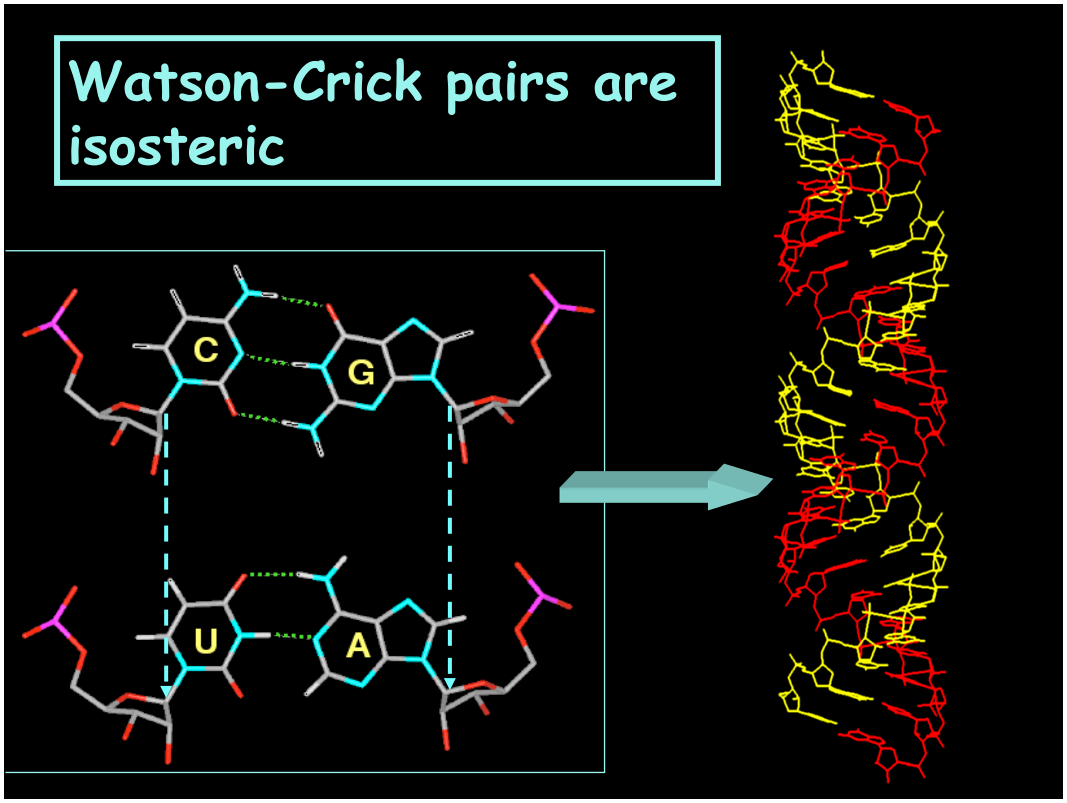


	J31		J23	J12
A Cases:10	2±1	<	6±3	2±2
B Cases:6	4±2	~	4±2	4±2
C Cases:17	6±3	>	3±2	2±3

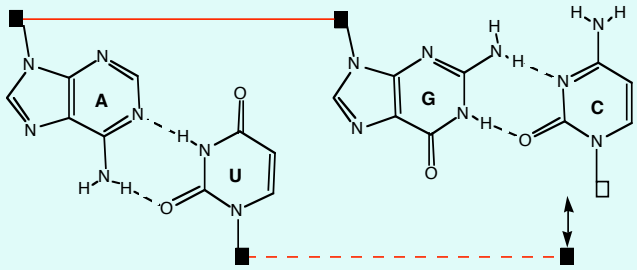




Which sequences are compatible with a given module?

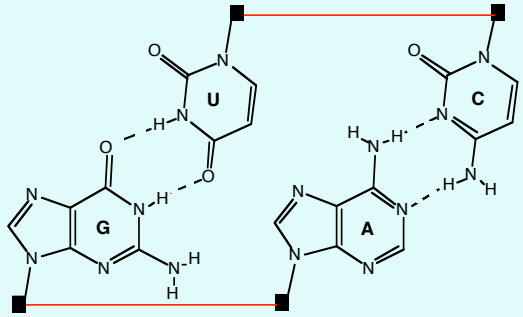


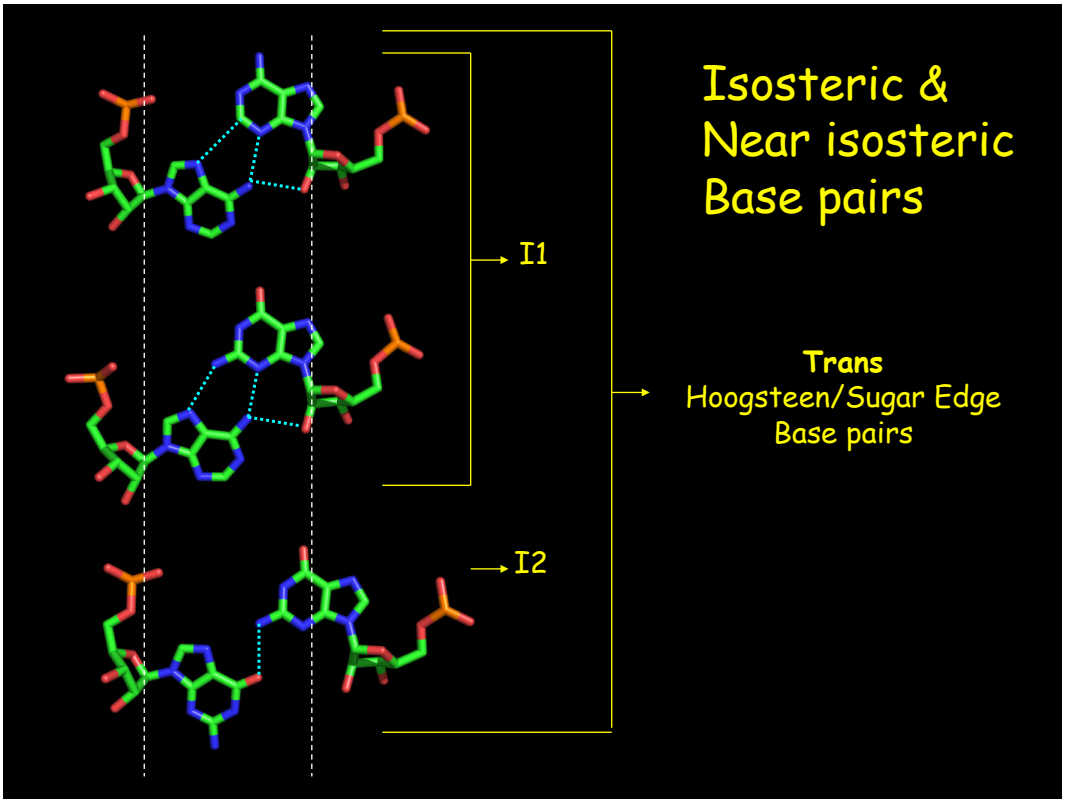
Non-Isosteric



Trans Watson-Crick/Watson-Crick

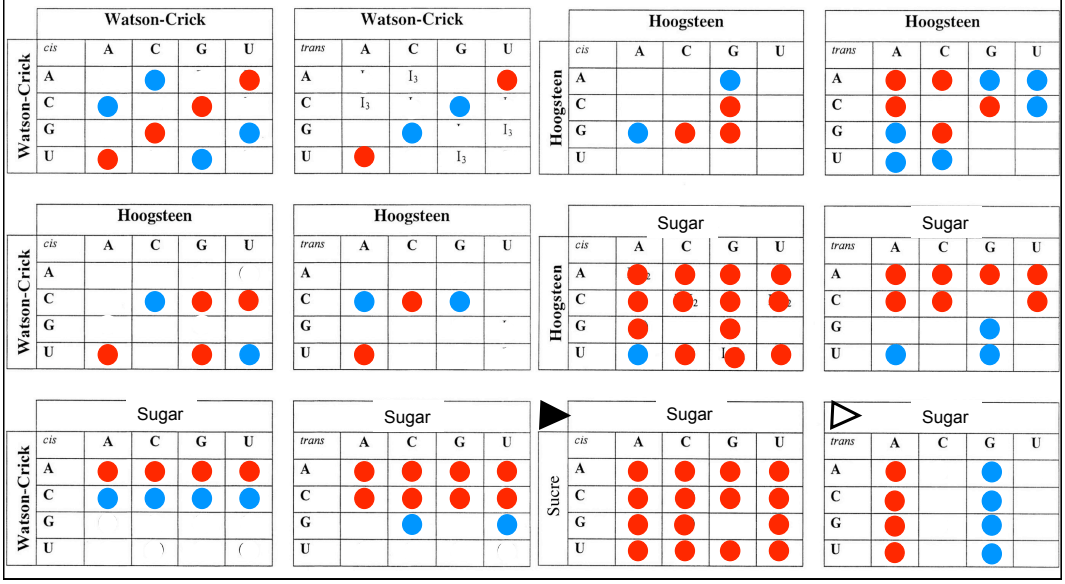
Isosteric





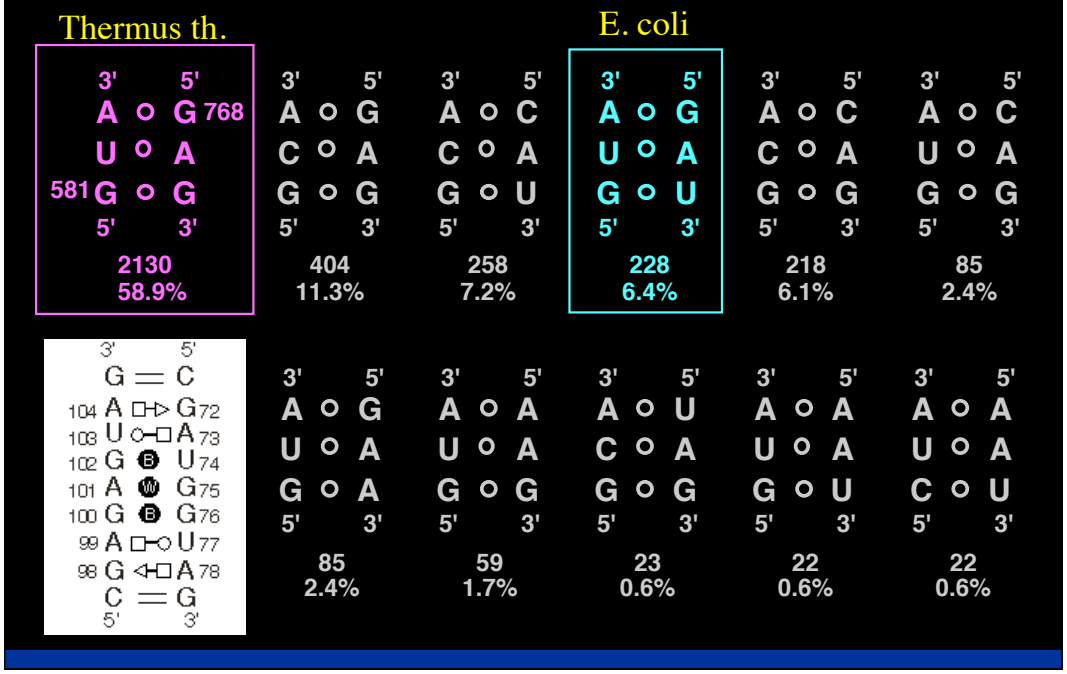
Isostericity Matrices

12 pairing families >> 12 isostericity matrices



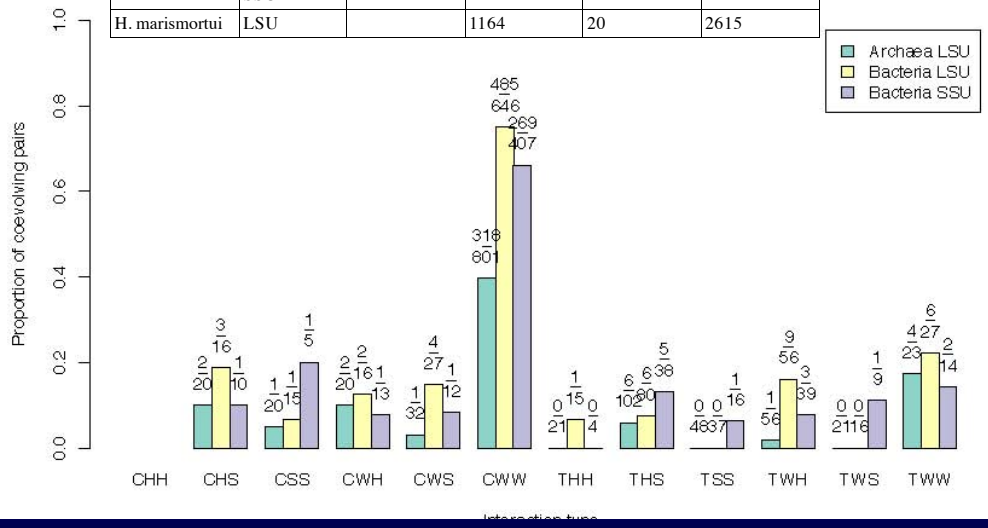
The module diversity is not simply the combinatorics of the isosteric base pairs

Loop E Motif in 16S rRNA (h20)



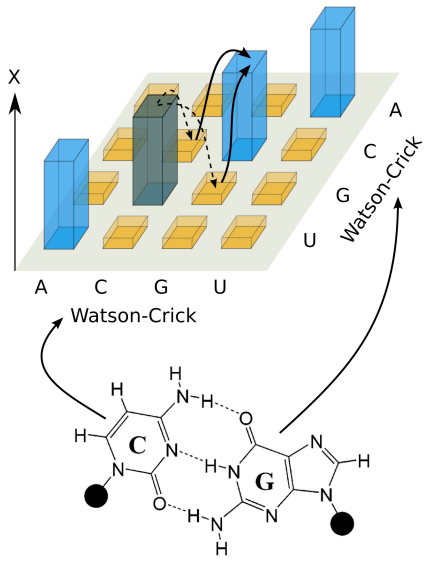
Are all base pairs evolving similarly ?

Species	SU	Structure	# interactions	# sequences	# sites
E. coli	LSU	[PDB ID]	951	214	2244
E. coli	SSU		567	175	1319
H. marismortui	LSU		1164	20	2615

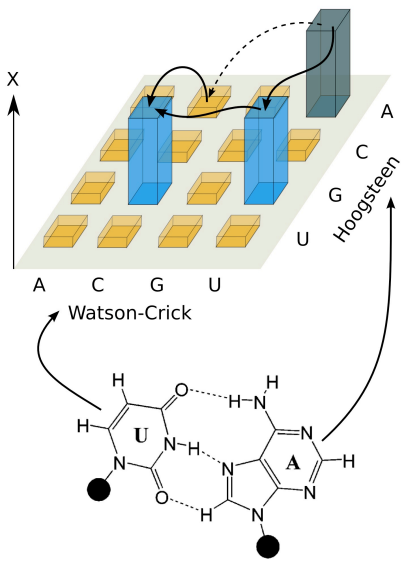


- Only $\sim$ 20% of non-WC interactions show a significant coevolutionary signal,
- Whereas more than 60% of WC pairs are coevolving.

In **coevolution**, a change from one state to the other implies a simultaneous substitution event at a different position.



Neutral networks connect different sequence states, distant by only one mutation, that possess the same fitness.

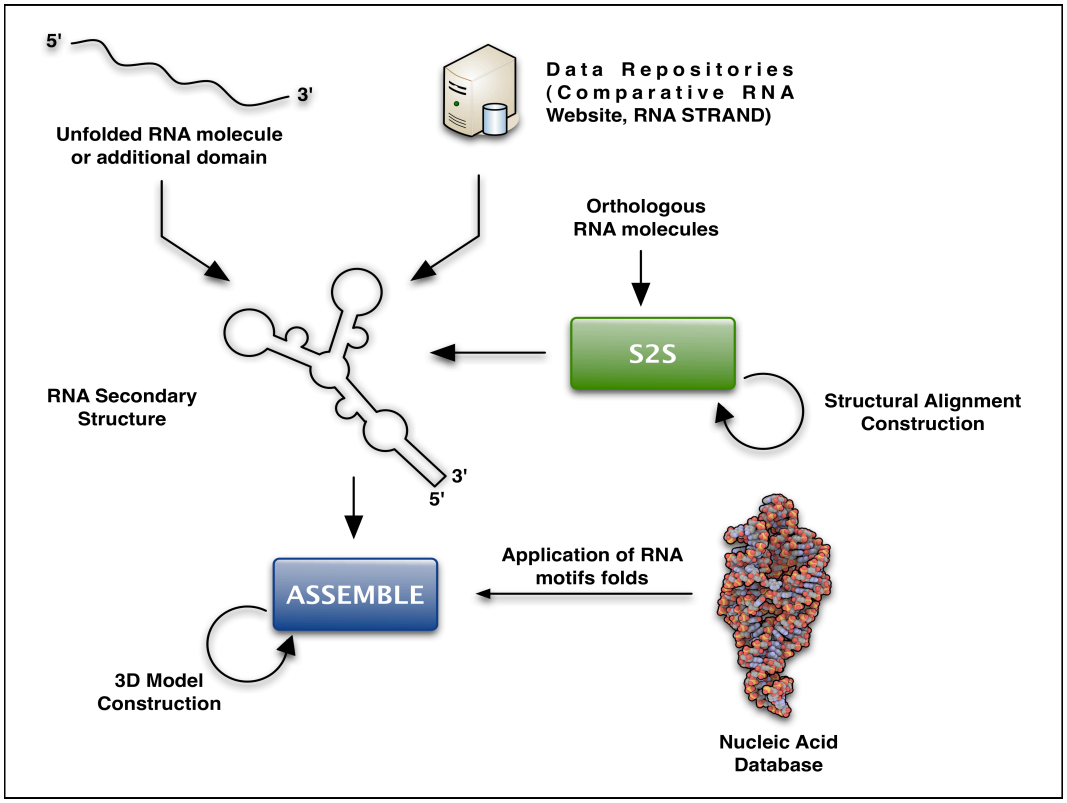


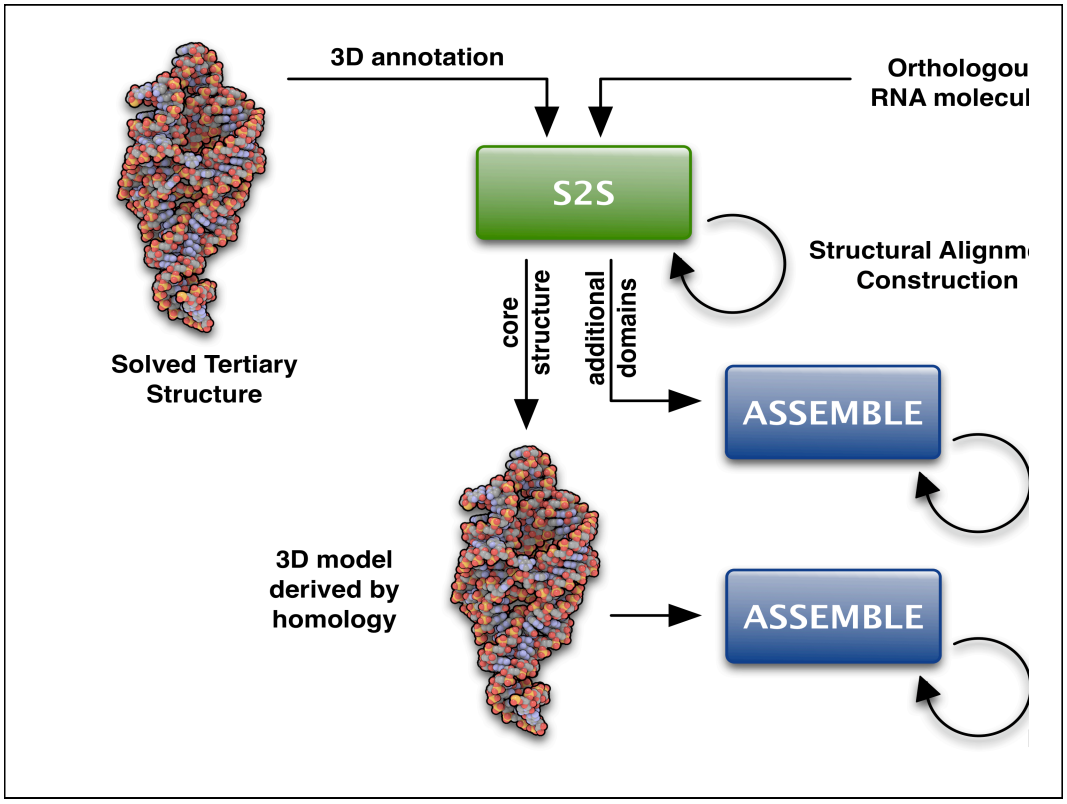
Conclusions

1. The cis-WW pairs have a very high substitution rate and the highest proportion of coevolving pairs.
2. The non-WC pairs display a lack of co-evolution signal.
3. Existence of neutral networks (based on isostericity matrices) decoupling the evolution at interacting sites.
4. Lead to lack of 3D signal in sequences.

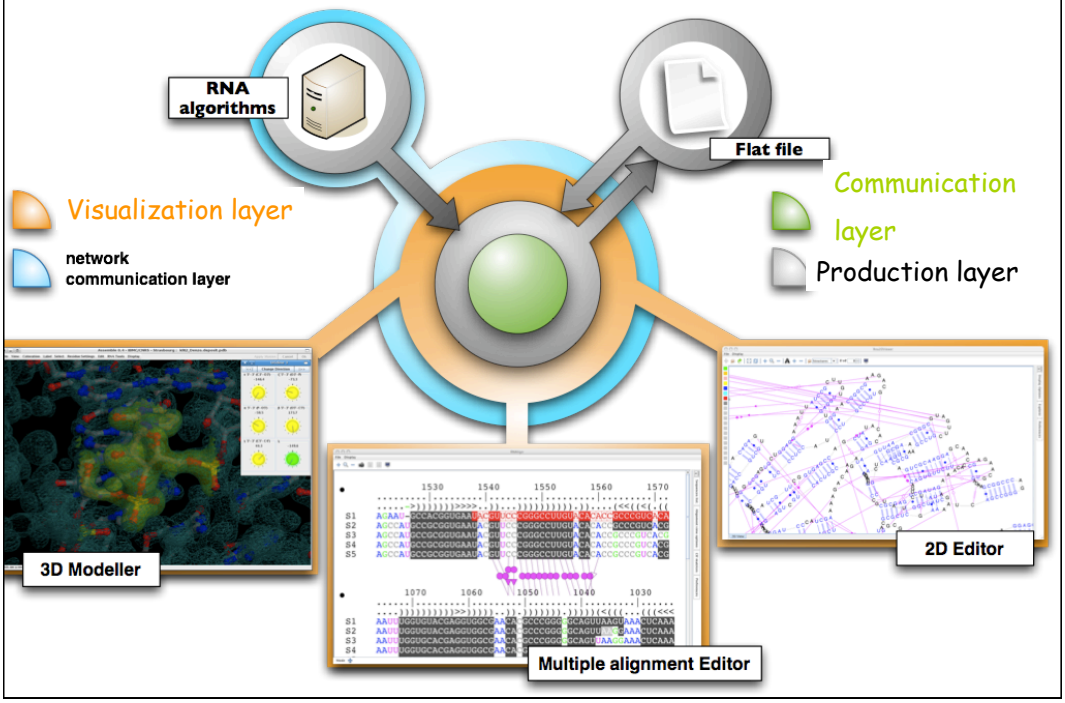
Modelling

by modular assembly


5'
3'
Unfolded RNA molecule
or additional domain
Data Repositories
(Comparative RNA
Website, RNA STRAND)
Orthologous
RNA molecules
S2S
Structural Alignment
Construction
RNA Secondary
Structure
3'
5'
ASSEMBLE
Application of RNA
motifs folds
3D Model
Construction
Nucleic Acid
Database


3D annotation
Orthologou
RNA molecul
S2S
Structural Alignm
Construction
core
structure
additional
domains
Solved Tertiary
Structure
ASSEMBLE
3D model
derived by
homology
ASSEMBLE

ASSEMBLE : from 1D to 2D and 3D



UPR 9002 du CNRS,
Architecture et Réactivité de l'ARN,
Institut de Biologie Moléculaire et Cellulaire,
Université de Strasbourg

Fabrice Jossinet, Benoît Masquida, José Cruz

Jiro Kondo, Pascal Auffinger, Quentin Vicens

Neocles Leontis, Jesse Stombaugh (OBGU)

Julien Dutheil

Assemble

http://bioinformatics.org/assemble

S2S

http://bioinformatics.org/s2s

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ DE STRASBOURG